Terrorism Hotspots

This data project has been used as an assignment during the LGM Data Science Virtual Internship.

Assignment

Imagine you are a security or defense analyst. Analyze the data and draw conclusions on the distribution and nature of terrorist incidents recorded around the world. In your analysis, include maps that visualize the location of different incidents. Your analysis may also provide answers to the following questions:

1. How has the number of terrorist activities changed over the years? Are there certain regions where this trend is different from the global averages?
2. Is the number of incidents and the number of casualties correlated? Can you spot any irregularities or outliers?
3. What are the most common methods of attacks? Does it differ in various regions or in time?
4. Plot the locations of attacks on a map to visualize their regional spread;

You are also free to explore the data further and extract additional insights other than the questions above.

Data Description

The provided compressed file globalterrorismdb_0718dist.tar.bz2 is an extract from the Global Terrorism Database (GTD) - an open-source database including information on terrorist attacks around the world from 1970 through 2017. The GTD includes systematic data on domestic as well as international terrorist incidents that have occurred during this time period and now includes more than 180,000 attacks. The database is maintained by researchers at the National Consortium for the Study of Terrorism and Responses to Terrorism (START), headquartered at the University of Maryland.

Since the number of variables and instances is very large, for this project, feel free to select a subset of columns or a specific timeframe.

Explanation of selected columns:

- Success - Success of a terrorist strike
- Suicide - 1 = "Yes" The incident was a suicide attack. 0 = "No" There is no indication that the incident was a suicide
- Attacktype1 - The general method of attack
- Attacktype1_txt - The general method of attack and broad class of tactics used
- Targtype1_txt - The general type of target/victim
- Targsubtype1_txt - The more specific target category
- Target1 - The specific person, building, installation that was targeted and/or victimized
- Natlty1_txt - The nationality of the target that was attacked
- Gname - The name of the group that carried out the attack
- Gsubname - Additional details about the group that carried out the attack like fractions

- Nperps - The total number of terrorists participating in the incident
- Weaptype1_txt - General type of weapon used in the incident
- Weapsubtype1_txt - More specific value for most of the Weapon Types
- Nkill - The number of total confirmed fatalities for the incident
- Nkillus - The number of U.S. citizens who died as a result of the incident

Libraries Used

In [28]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings("ignore")
```

Read the Data

In [38]:
```
df_data = pd.read_csv("globalterrorismdb_0718dist.tar.bz2", compression

# Display the DataFrame without the first column.
df_data.head()
```

Out[38]:

	eventid	iyear	imonth	iday	approxdate	extended	resolution	country	country_txt
0	197000000001	1970	7	2	NaN	0	NaN	58	Dominican Republic
1	197000000002	1970	0	0	NaN	0	NaN	130	Mexico
2	197001000001	1970	1	0	NaN	0	NaN	160	Philippines
3	197001000002	1970	1	0	NaN	0	NaN	78	Greece
4	197001000003	1970	1	0	NaN	0	NaN	101	Japan

5 rows × 135 columns

In [39]:
```
df_data.shape
```

Out[39]:
```
(181691, 135)
```

In [81]:
```
df_data.describe()
```

Out[81]:

	eventid	iyear	imonth	iday	extended	cou
count	1.816910e+05	181691.000000	181691.000000	181691.000000	181691.000000	181691.00
mean	2.002705e+11	2002.638997	6.467277	15.505644	0.045346	131.96
std	1.325957e+09	13.259430	3.388303	8.814045	0.208063	112.41
min	1.970000e+11	1970.000000	0.000000	0.000000	0.000000	4.00
25%	1.991021e+11	1991.000000	4.000000	8.000000	0.000000	78.00
50%	2.009022e+11	2009.000000	6.000000	15.000000	0.000000	98.00
75%	2.014081e+11	2014.000000	9.000000	23.000000	0.000000	160.00
max	2.017123e+11	2017.000000	12.000000	31.000000	1.000000	1004.00

8 rows × 77 columns

In [84]:
```
df_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 181691 entries, 0 to 181690
Columns: 135 entries, eventid to related
dtypes: float64(55), int64(22), object(58)
memory usage: 187.1+ MB
```

In [46]:
```
for i in df_data:
    print(i,end = ', ')
```

```
eventid, iyear, imonth, iday, approxdate, extended, resolution, country, c
ountry_txt, region, region_txt, provstate, city, latitude, longitude, spec
ificity, vicinity, location, summary, crit1, crit2, crit3, doubtterr, alte
rnative, alternative_txt, multiple, success, suicide, attacktype1, attackt
ype1_txt, attacktype2, attacktype2_txt, attacktype3, attacktype3_txt, targ
type1, targtype1_txt, targsubtype1, targsubtype1_txt, corp1, target1, natl
ty1, natlty1_txt, targtype2, targtype2_txt, targsubtype2, targsubtype2_tx
t, corp2, target2, natlty2, natlty2_txt, targtype3, targtype3_txt, targsub
type3, targsubtype3_txt, corp3, target3, natlty3, natlty3_txt, gname, gsub
name, gname2, gsubname2, gname3, gsubname3, motive, guncertain1, guncertai
n2, guncertain3, individual, nperps, nperpcap, claimed, claimmode, claimmo
de_txt, claim2, claimmode2, claimmode2_txt, claim3, claimmode3, claimmode3
_txt, compclaim, weaptype1, weaptype1_txt, weapsubtype1, weapsubtype1_txt,
weaptype2, weaptype2_txt, weapsubtype2, weapsubtype2_txt, weaptype3, weapt
ype3_txt, weapsubtype3, weapsubtype3_txt, weaptype4, weaptype4_txt, weapsu
btype4, weapsubtype4_txt, weapdetail, nkill, nkillus, nkillter, nwound, nw
oundus, nwoundte, property, propextent, propextent_txt, propvalue, propcom
ment, ishostkid, nhostkid, nhostkidus, nhours, ndays, divert, kidhijcountr
y, ransom, ransomamt, ransomamtus, ransompaid, ransompaidus, ransomnote, h
ostkidoutcome, hostkidoutcome_txt, nreleased, addnotes, scite1, scite2, sc
ite3, dbsource, INT_LOG, INT_IDEO, INT_MISC, INT_ANY, related,
```

In [98]:
```
df = df_data[['success','suicide','iyear','region','region_txt' ,'event
df
```

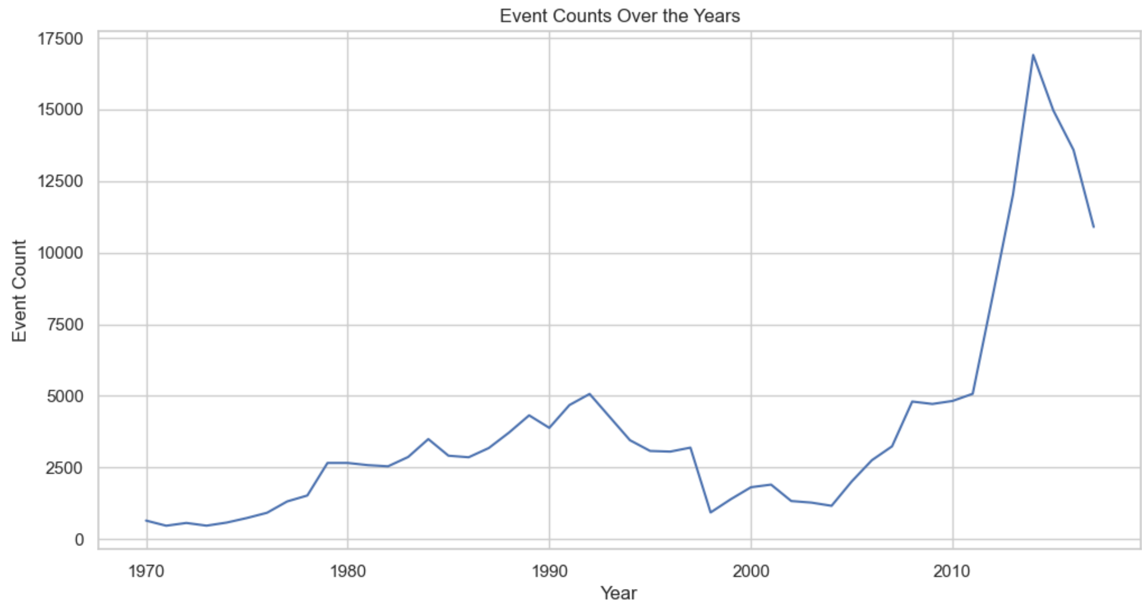
Out[98]:

	success	suicide	iyear	region	region_txt	eventid	attacktype1	attacktype
0	1	0	1970	2	Central America & Caribbean	197000000001	1	Assassir
1	1	0	1970	1	North America	197000000002	6	Hostage T (Kidnap
2	1	0	1970	5	Southeast Asia	197001000001	1	Assassir
3	1	0	1970	8	Western Europe	197001000002	3	Bombing/Expl
4	1	0	1970	4	East Asia	197001000003	7	Facility/Infrastru /
...	...	...	...	...	...	...	...	
181686	1	0	2017	11	Sub-Saharan Africa	201712310022	2	Armed As
181687	1	0	2017	10	Middle East & North Africa	201712310029	3	Bombing/Expl
181688	1	0	2017	5	Southeast Asia	201712310030	7	Facility/Infrastru /
181689	0	0	2017	6	South Asia	201712310031	3	Bombing/Expl
181690	0	0	2017	5	Southeast Asia	201712310032	3	Bombing/Expl

181691 rows × 19 columns

1.How has the number of terrorist activities changed over the years? Are there certain regions where this trend is different from the global averages?

In [56]:
```
event_counts = df.groupby(['iyear'])['eventid'].count().reset_index()

# Create a line plot using Seaborn
sns.set(style="whitegrid")
plt.figure(figsize=(12, 6))
sns.lineplot(x='iyear', y='eventid', data=event_counts)

# Set labels and title
plt.xlabel('Year')
plt.ylabel('Event Count')
plt.title('Event Counts Over the Years')

# Show the plot
plt.show()
```



In [60]:
```
region_over_years = pd.crosstab(df.iyear,df.region_txt)
region_over_years
```

Out[60]:

region_txt iyear	Australasia & Oceania	Central America & Caribbean	Central Asia	East Asia	Eastern Europe	Middle East & North Africa	North America	South America	Sout Asi
1970	1	7	0	2	12	28	472	65	
1971	1	5	0	1	5	55	247	24	
1972	8	3	0	0	1	53	73	33	
1973	1	6	0	2	1	19	64	83	
1974	1	11	0	4	2	42	111	81	
1975	0	9	0	12	0	44	159	55	
1976	0	45	0	2	0	55	125	91	
1977	0	24	0	4	2	211	149	119	

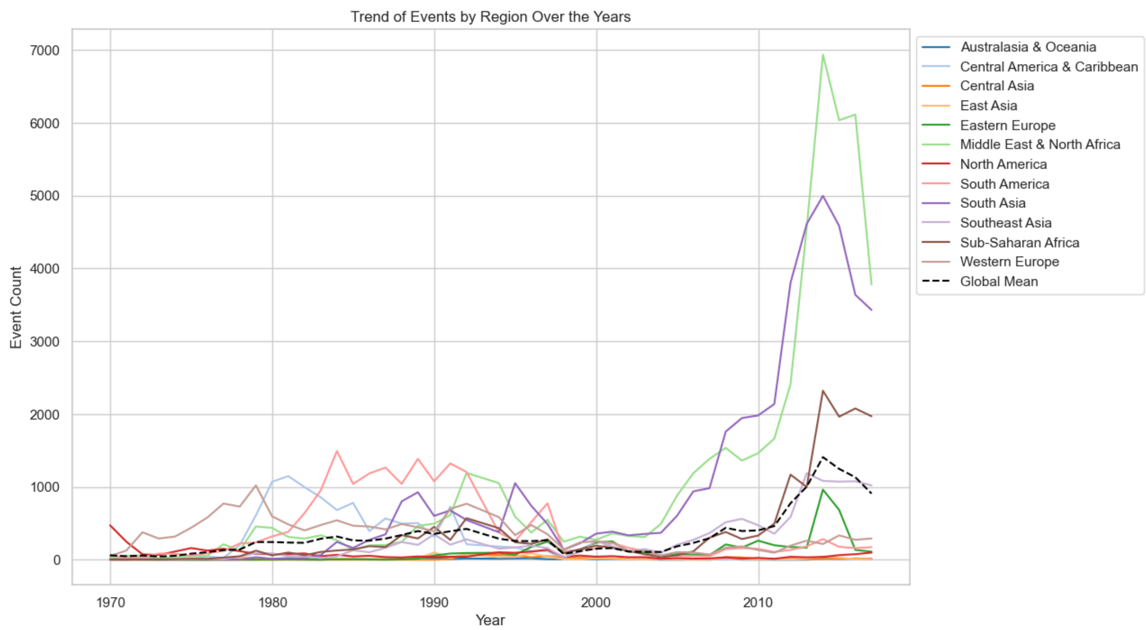
In [91]:

```
# Group the data by 'iyear' and 'region_txt' and count the occurrences 
global_count_by_year_and_region = df.groupby(['iyear', 'region_txt'])['

# Calculate the average count for each year
average_count_by_year = global_count_by_year_and_region.groupby('iyear'

# Print the resulting DataFrame
print(average_count_by_year)
```

...

In [95]:

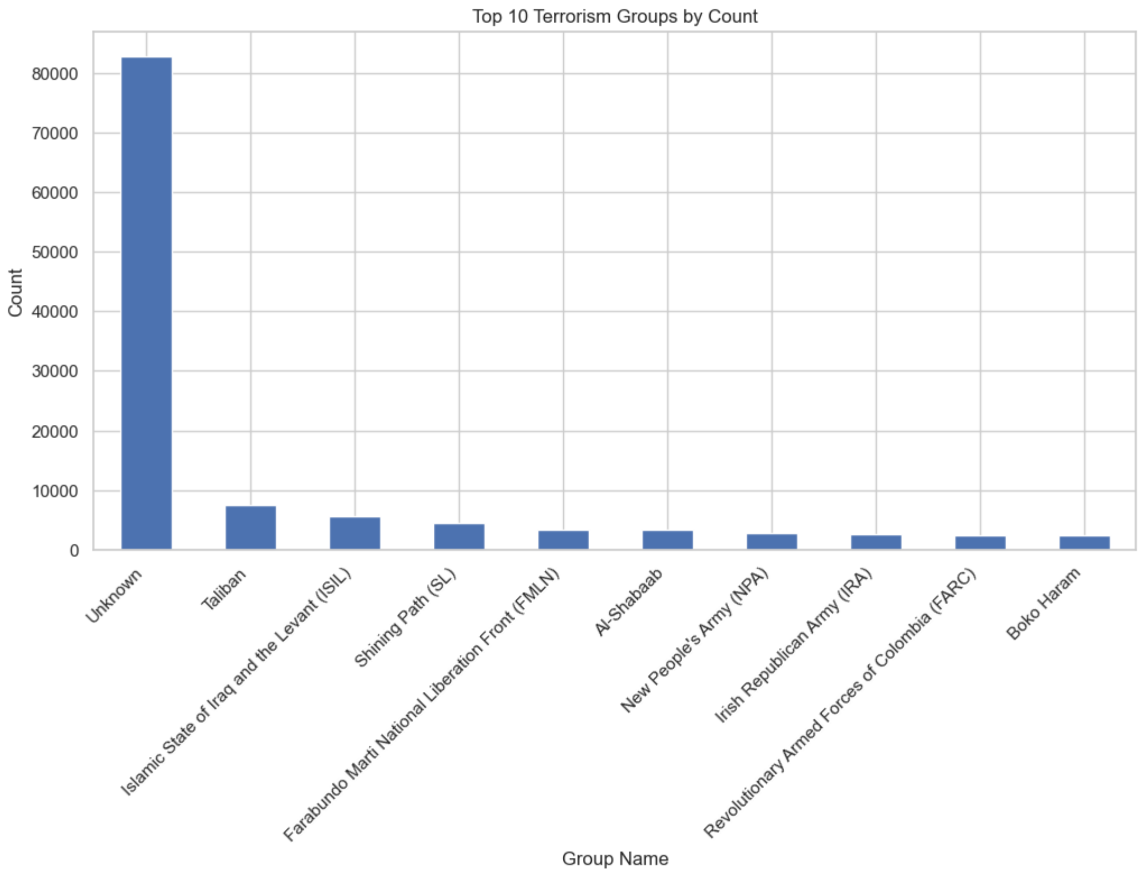
```
# Set the figure size and style
plt.figure(figsize=(12, 8))
sns.set_style("whitegrid")

# Define a list of regions to loop through
regions = ['Australasia & Oceania', 'Central America & Caribbean', 'Cen
           'Eastern Europe', 'Middle East & North Africa', 'North Ameri
           'South Asia', 'Southeast Asia', 'Sub-Saharan Africa', 'Weste

# Create a color palette for the regions
colors = sns.color_palette("tab20", len(regions))

# Create a trend chart using Seaborn lineplot for each region
for i, region in enumerate(regions):
    sns.lineplot(data=region_over_years, x='iyear', y=region, label=reg

# Add the global mean as a lineplot
sns.lineplot(data=average_count_by_year, x='iyear', y='eventid', label=

# Set labels and title
plt.xlabel('Year')
plt.ylabel('Event Count')
plt.title('Trend of Events by Region Over the Years')

# Customize the legend
plt.legend(loc='upper left', bbox_to_anchor=(1, 1))

# Show the plot
plt.show()

```



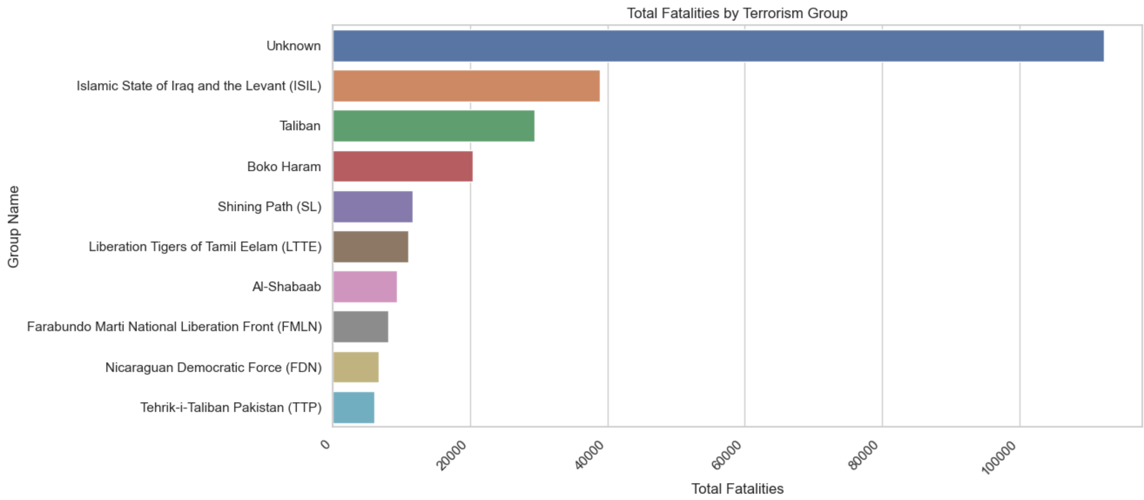
2. Is the number of incidents and the number of casualties correlated? Can you spot any irregularities or outliers?

In [115]:

```
top_10_groups = df['gname'].value_counts().head(10)

# Create a bar plot for the top 10 'gname' values
plt.figure(figsize=(12, 6))
top_10_groups.plot(kind='bar')
plt.title('Top 10 Terrorism Groups by Count')
plt.xlabel('Group Name')
plt.ylabel('Count')
plt.xticks(rotation=45, ha='right')
plt.show()
```



In [125]:
```python
Top10_Gname = df.groupby('gname')['nkill'].sum()

# Sort the result in ascending order by the total sum of 'nkill'
sorted_data = Top10_Gname.sort_values(ascending=False)[:10]

# Create a bar plot for the total sum of 'nkill' for each group using S
plt.figure(figsize=(12, 6))
sns.barplot(x=sorted_data.values, y=sorted_data.index, orient='h')
plt.title('Total Fatalities by Terrorism Group')
plt.xlabel('Total Fatalities')
plt.ylabel('Group Name')
plt.xticks(rotation=45, ha='right')
plt.show()
```

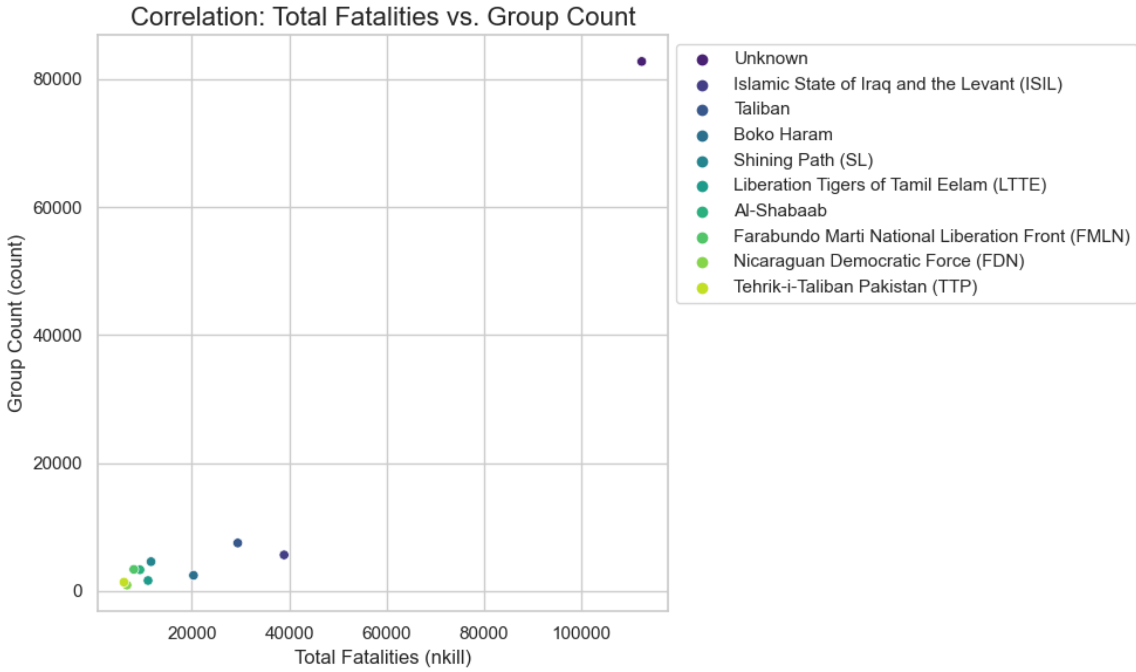


In [137]:
```python
total_fatalities = df.groupby('gname')['nkill'].sum()

# Sort the result in ascending order by the total sum of 'nkill' and ta
top_10_groups = total_fatalities.sort_values(ascending=False).head(10)

# Calculate the count of occurrences of each group
group_counts = df['gname'].value_counts().reset_index()
group_counts.columns = ['gname', 'count']

# Merge the two DataFrames based on the 'gname' column and include only
merged_data = pd.merge(top_10_groups, group_counts, on='gname')

# Print the resulting merged DataFrame
print(merged_data)
```

```
                                             gname     nkill  count
0                                          Unknown  112367.0  82782
1      Islamic State of Iraq and the Levant (ISIL)   38923.0   5613
2                                          Taliban   29410.0   7478
3                                       Boko Haram   20328.0   2418
4                                Shining Path (SL)   11601.0   4555
5          Liberation Tigers of Tamil Eelam (LTTE)   10989.0   1606
6                                       Al-Shabaab    9330.0   3288
7  Farabundo Marti National Liberation Front (FMLN)    8065.0   3351
8                Nicaraguan Democratic Force (FDN)    6662.0    895
9                   Tehrik-i-Taliban Pakistan (TTP)    6042.0   1351
```

In [143]:

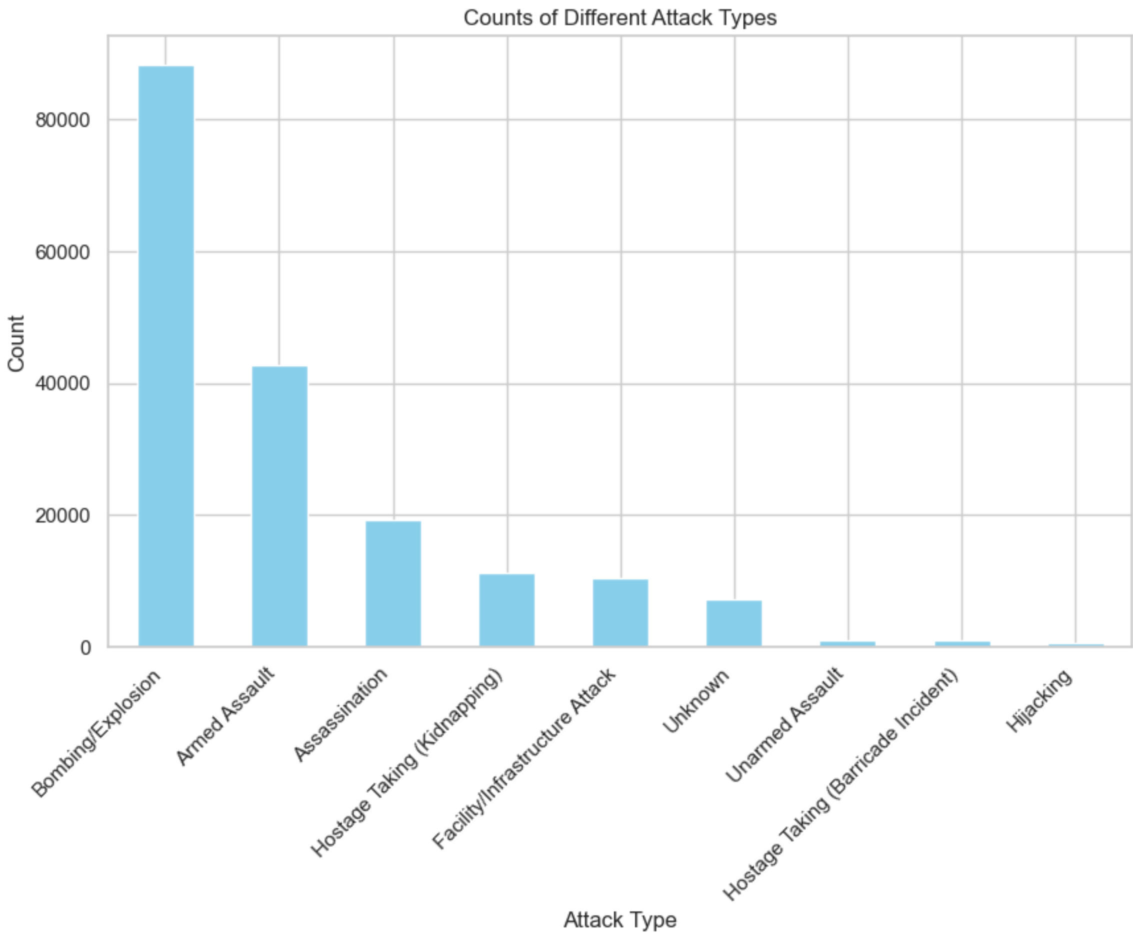
```python
plt.figure(figsize=(10, 6))  # Adjust figure size
sns.set_style("whitegrid")  # Set a white grid background

# Create the scatter plot
scatter = sns.scatterplot(data=merged_data, x='nkill', y='count', hue='

# Customize the title and labels
plt.title('Correlation: Total Fatalities vs. Group Count', fontsize=16)
plt.xlabel('Total Fatalities (nkill)', fontsize=12)
plt.ylabel('Group Count (count)', fontsize=12)

# Adjust the legend
scatter.legend(title='Group Name', title_fontsize=12)
plt.legend(loc='upper left', bbox_to_anchor=(1, 1))

# Show the plot
plt.tight_layout()
plt.show()
```



3. What are the most common methods of attacks? Does it differ in various regions or in time?

In [154]:

```
attack_type_counts = df['attacktype1_txt'].value_counts()

# Create a bar plot
plt.figure(figsize=(10, 6))
attack_type_counts.plot(kind='bar', color='skyblue')
plt.title('Counts of Different Attack Types')
plt.xlabel('Attack Type')
plt.ylabel('Count')
plt.xticks(rotation=45, ha='right')
plt.show()
```

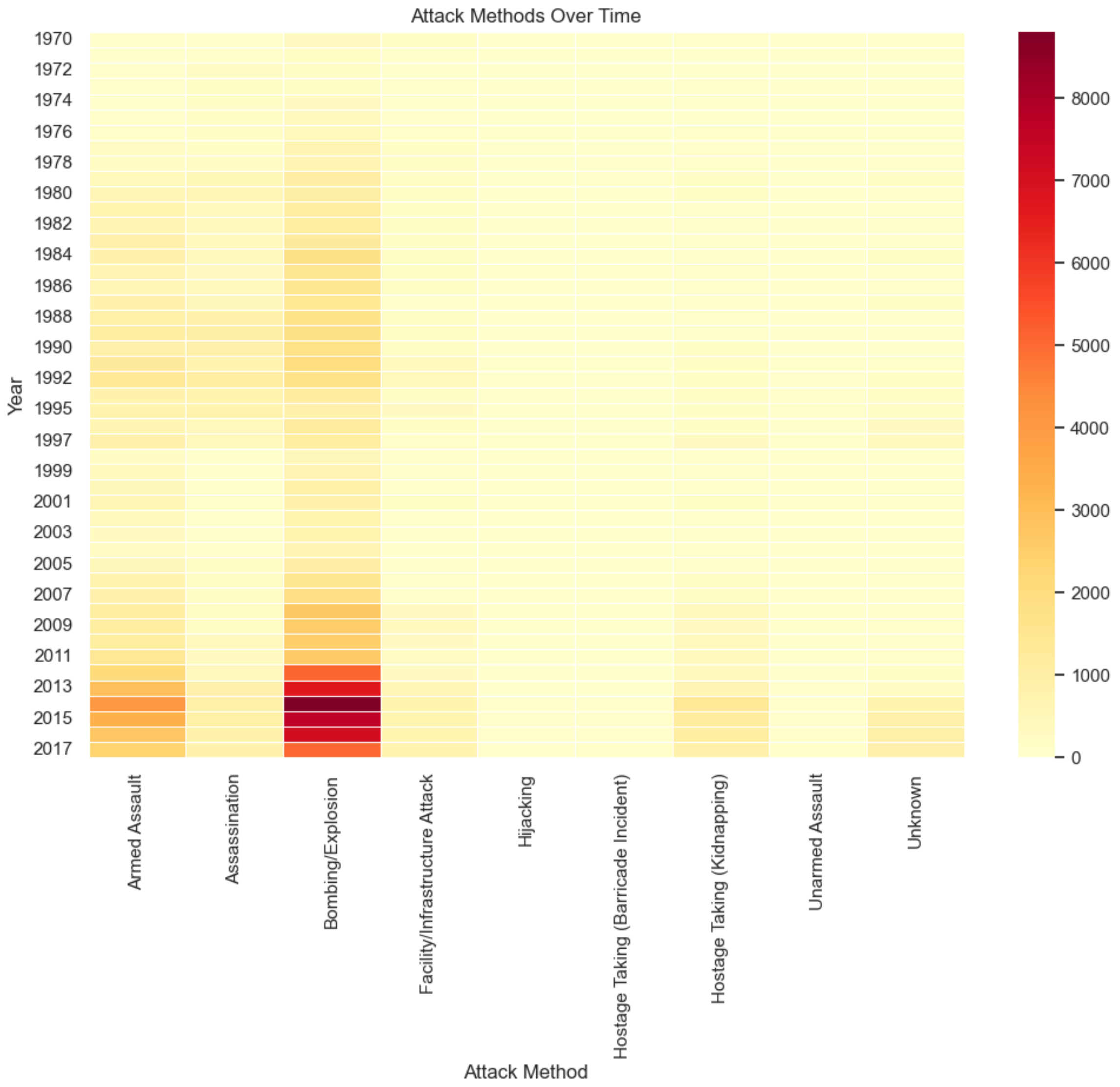


In [152]:
```
common_methods_by_region = df.groupby('region_txt')['attacktype1_txt'].

# Print the most common methods in each region
print("\nMost common attack methods by region:")
print(common_methods_by_region)
```

```
Most common attack methods by region:
region_txt                   attacktype1_txt
Australasia & Oceania        Bombing/Explosion      75
Central America & Caribbean  Armed Assault        4361
Central Asia                 Bombing/Explosion     235
East Asia                    Bombing/Explosion     330
Eastern Europe               Bombing/Explosion    2766
Middle East & North Africa   Bombing/Explosion   30908
North America                Bombing/Explosion    1534
South America                Bombing/Explosion    9039
South Asia                   Bombing/Explosion   21246
Southeast Asia               Bombing/Explosion    4818
Sub-Saharan Africa           Armed Assault        6004
Western Europe               Bombing/Explosion    8508
Name: attacktype1_txt, dtype: int64
```

In [153]:

```
attack_methods_over_time = df.groupby(['iyear', 'attacktype1_txt']).siz

# Plot the trends using a heatmap
plt.figure(figsize=(12, 8))
sns.heatmap(attack_methods_over_time, cmap='YlOrRd', linewidths=0.5)
plt.title('Attack Methods Over Time')
plt.xlabel('Attack Method')
plt.ylabel('Year')
plt.show()
```



4. Plot the locations of attacks on a map to visualize their regional spread;

In [167]:
```
df.groupby('natlty1_txt')['gname'].value_counts()
```

Out[167]:
```
natlty1_txt  gname
Afghanistan  Taliban                                    6565
             Unknown                                    3942
             Khorasan Chapter of the Islamic State       225
             Haqqani Network                              49
             Hizb-I-Islami                                29
                                                        ...
Zimbabwe     Guerrillas                                    1
             Gunmen                                        1
             Liberation War Veterans Association           1
             National Youth Service of Zimbabwe            1
             South African guerrillas                      1
Name: gname, Length: 6801, dtype: int64
```